Raise a Child in Large Language Model: Towards Effective and Generalizable Fine-tuning

Year: 2021 | Citations: 212 | Authors: Runxin Xu, Fuli Luo, Zhiyuan Zhang

---

Abstract

Recent pretrained language models extend from millions to billions of parameters. Thus the need to fine-tune an extremely large pretrained model with a limited training corpus arises in various downstream tasks. In this paper, we propose a straightforward yet effective fine-tuning technique, Child-Tuning, which updates a subset of parameters (called child network) of large pretrained models via strategically masking out the gradients of the non-child network during the backward process. Experiments on various downstream tasks in GLUE benchmark show that Child-Tuning consistently outperforms the vanilla fine-tuning by 1.5 8.6 average score among four different pretrained models, and surpasses the prior fine-tuning techniques by 0.6 1.3 points. Furthermore, empirical results on domain transfer and task transfer show that Child-Tuning can obtain better generalization performance by large margins.